Comparison of eight auditory models of normal-hearing processing

Alejandro Osses

E-mail: ale.a.osses@gmail.com

Laboratoire des Systèmes Perceptifs, École Normale Supérieure (ENS), PSL University, Paris, France

31 May 2022

WASdag, Utrecht, the Netherlands

Acta Acustica 2022, **6**, 17
© The Author(s), Published by EDP Sciences, 2022
https://doi.org/10.1051/aacus/2022008

Available online at:
https://acta-acustica.edpsciences.org

Topical Issue - Auditory models: from binaural processing to multimodal cognition

SCIENTIFIC ARTICLE OPEN ACCESS

A comparative study of eight human auditory models of monaural processing

Alejandro Osses Vecchi[1,*], Léo Varnet[1], Laurel H. Carney[2], Torsten Dau[3], Ian C. Bruce[4], Sarah Verhulst[5], and Piotr Majdak[6]

[1] Laboratoire des systè... ...ptifs, Département d'études cognitives, École Normale Supérieure, PSL University, CNRS, 75005 Paris, ...
[2] Departments of Biomedica... ...cience, University of Rochester, Rochester, NY 14642, USA
[3] Hearing Systems Section, Departme... ...Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark
[4] Department of Electrical and Computer Engine... ...ity, Hamilton, ON L8S 4K1, Canada
[5] Hearing Technology group, WAVES, Department of Info... ...t University, 9000 Ghent, Belgium
[6] Acoustics Research Institute, Austrian Academy of Sciences, 1040 V...

Many thanks to all my collaborators!

Received 4 July 2021, Accepted 28 February 2022

Abstract – A number of auditory models have been developed using diverging approaches, either physiological or perceptual, but they share comparable stages of signal processing, as they are inspired by the same constitutive parts of the auditory system. We compare eight monaural models that are openly accessible in the Auditory Modelling Toolbox. We discuss the considerations required to make the model outputs comparable to each other, as well as the results for the following model processing stages or their equivalents: Outer and middle ear, cochlear filter bank, inner hair cell, auditory nerve synapse, cochlear nucleus, and inferior colliculus. The discussion includes a list of recommendations for future applications of auditory models.

My background in the “last years”

2014-2018 TU Eindhoven – Ph.D. in Auditory modelling
(prof. Armin Kohlrausch)

2018-2020 UGent – PostDoc in Auditory modelling
(prof. Sarah Verhulst, prof. Dick Botteldooren)

2021 – ENS : **fastACI** project with Léo Varnet

2020 – Developer within the Auditory Modelling Toolbox project
amtoolbox.sourceforge.net (lead by dr. Piotr Majdak)

Motivation (1 / 2)

Based on:

Osses, Varnet, Carney, Dau, Bruce, Verhulst, Majdak (2022, Acta Acustica)
Open Access, available at:
https://acta-acustica.edpsciences.org/articles/aacus/abs/2022/01/aacus210060/aacus210060.html

Abstract:

A number of **auditory models** have been developed using diverging approaches, either **physiological** or **perceptual**, but they share comparable stages of signal processing, as they are inspired by the same constitutive parts of the auditory system. In this seminar, I will briefly describe the main stages of sound processing from the outer ear (or pinna) up to the inferior colliculus (midbrain) but I will focus on the physiological aspects that have been implemented in the model stages of **inner hair cell** (IHC) processing and **auditory nerve** (AN) **synapse**, at the beginning of the auditory **neural pathway**. I will also show auditory responses obtained from perceptual models that can capture specific neural processing properties.

Motivation (2 / 2)

From Maxwell et al. (2020, JASA):

Model by Zilany et al. (2014)

Note that this AN model is substantially different from peripheral models such as Dau *et al.* (1997), used in simulations of the notched-noise task in Derleth and Dau (2000) and similar masking tasks (with some changes to the model) in Jepsen *et al.* (2008). The Dau *et al.* (1997) model, and later modifications, did not include IHC saturation and did not aim to simulate physiological neural fluctuations and capture.

Another model

True

True, but this does not mean that a model may not account for neural fluctuations and capture

They investigated this phenomenon

"Publish your code: it is good enough" (Barnes, 2010, Nature)

So, if a code is published, you can try it!

What is a model?

https://en.wikipedia.org/wiki/Model

"A model is an informative representation of an object, person or system"

Physical model [edit]

A physical model is a physical representation of an object, such as a miniature aeroplane representing a real aeroplane.

- ~~Model (art),~~ a person posing for an artist, e.g. a 15th-century criminal representing the biblical Judas in Leonardo da Vinci's painting *The Last Supper*
- ~~Model (person),~~ a person who serves as a template for others to copy, often in the context of advertising commercial products
- ~~Model (product),~~ a particular design of a product offered by its manufacturer
- ~~Car model,~~ a particular design of vehicle sold by a manufacturer
- Model organism (often shortened to *model*), a non-human species that is studied to understand biological phenomena present in other related organisms, e.g. a guinea pig starved of vitamin C to study scurvy, an experiment that would be immoral to conduct on a person
- ~~Model (mimicry),~~ a species that is mimicked by another species

Conceptual model [edit]

A conceptual model is a theoretical representation of a system, e.g. a set of equations attempting to describe the workings of the atmosphere for the purpose of weather forecasting.

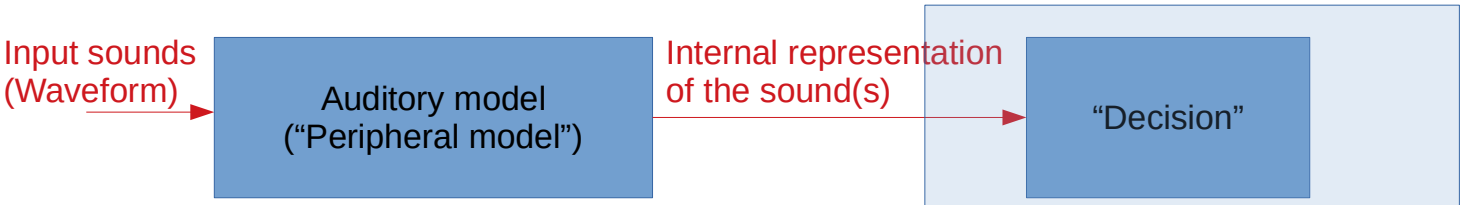
- Conceptual model (computer science), a representation of entities and their relationships
- Mathematical model, a description of a system using mathematical concepts and language
- ~~Economic model,~~ a theoretical construct representing economic processes
- Statistical model, a mathematical model that usually specifies the relationship between one or more random variables and other non-random variables
- ~~Model (CGI), a~~ mathematical representation of any surface of an object in three dimensions via specialized software
- Model (logic), a set along with a collection of finitary operations, and relations that are defined on it, satisfying a given collection of axioms
- ~~Model (MVC),~~ the central component of the model–view–controller software design pattern
- Standard model (disambiguation)
- ~~Medical model,~~ a proposed "set of procedures in which all doctors are trained"
- ~~Model act, a law~~ drafted centrally to be disseminated and proposed for enactment in multiple independent legislatures

What is a computational model?

Computational modelling is the use of computers to simulate and study complex systems using mathematics, physics and computer science.

(https://www.nibib.nih.gov/science-education/science-topics/computational-modeling)

I will focus on computational models of the auditory system



Simple task:
Yes, the target sound is present
No, the target sound is absent

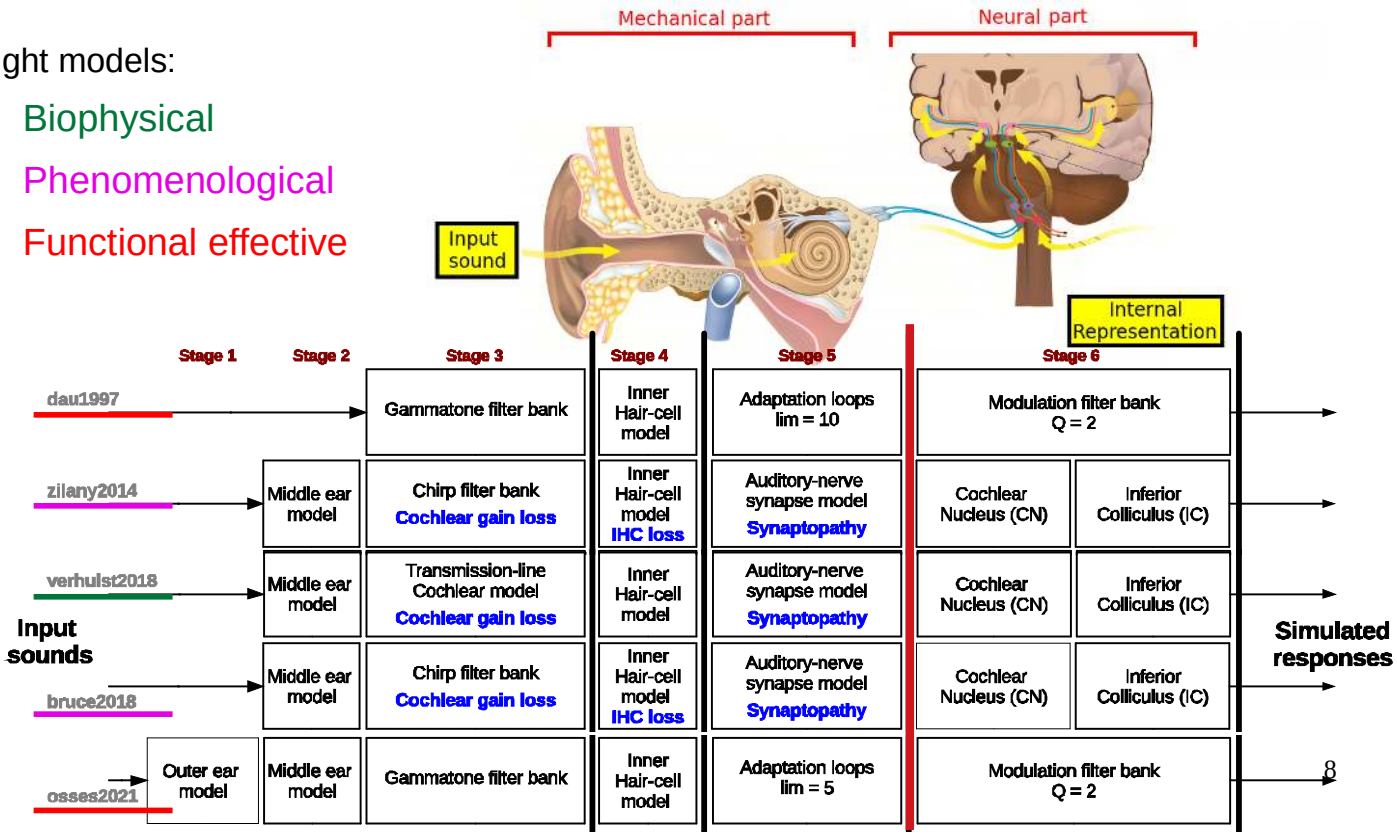
Biophysical model
Phenomenological model
Functional effective model
Statistical model

Not today, but see, e.g.,

Dau et al. (1997), Osses & Kohlrausch (2021)

Eight models:

- Biophysical
- Phenomenological
- Functional effective



Differences between the models

- They were built using different rationales:
  - Physiologists don't like functional effective (perceptual) models
  - Psychoacousticians do not always like the complexity of the biophysical and phenomenological models
  - Physiologists and psychoacousticians do not always use the same input stimuli in their model design
- **But in the end, all models are just an approximation…** (Majdak et al. 2022)
- **At the same time, the comparison among models should be "fair"**

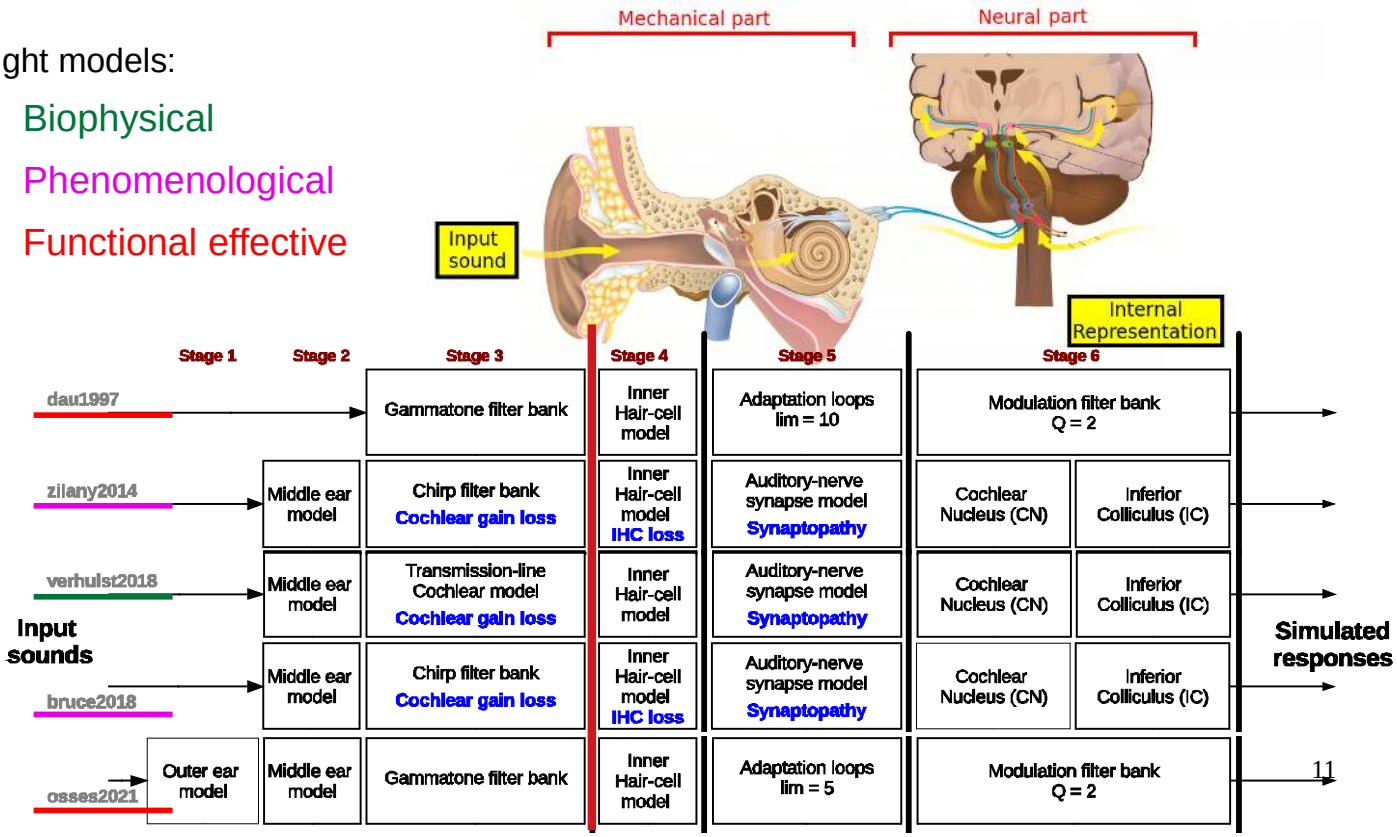
In the paper...

- **A common dataset of input sounds used in all models**

- Level conventions in each model
- Outer ear / middle ear: yes or no
- In the cochlear filter bank: **Identical CFs**, number of filters?
- IHC: Type of implementation
- Auditory Nerve synapse: Number of nerve fibres, PSTHs or mean rates?
- CN / IC: one modulation filter with the same centre frequency

Eight models:

- Biophysical
- Phenomenological
- Functional effective



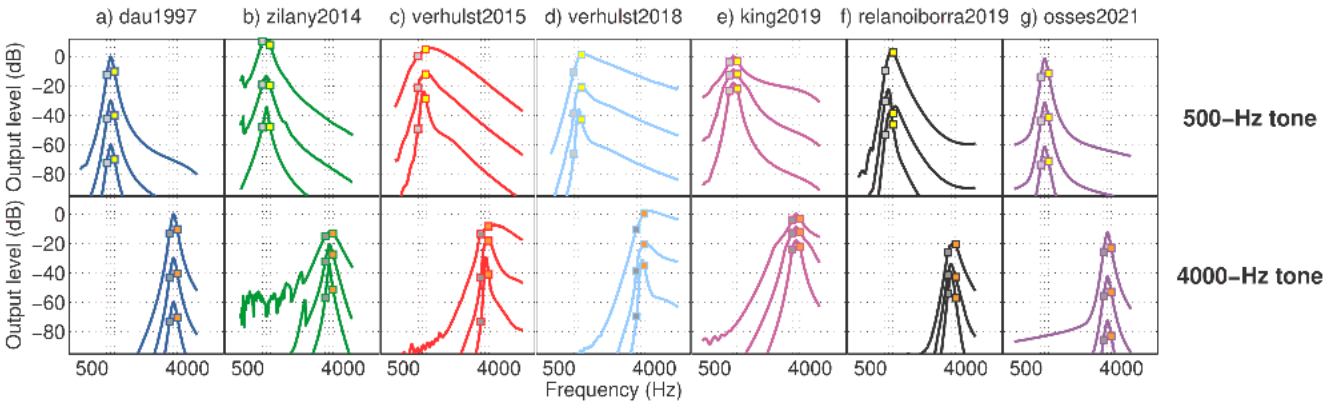
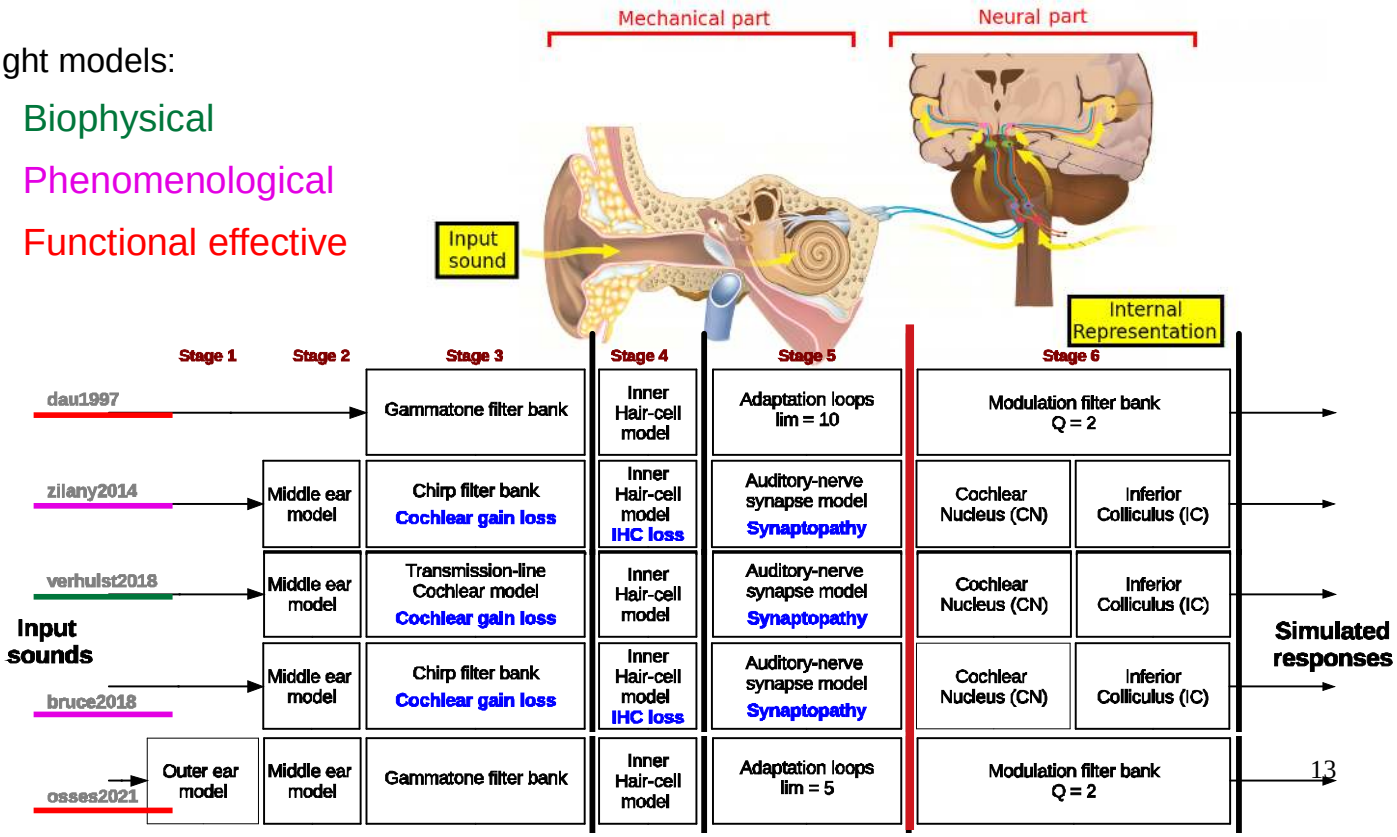


Figure 4. Filter bank responses to pure tones at 500 Hz and 4000 Hz (top two rows) and spectral magnitudes of single-channel responses to white noises (bottom two rows) for sounds of 40-, 70-, or 100-dB SPL (bottom-to-top coloured curves, respectively). In the first two rows, the responses represent excitation patterns at the simulated CFs. The coloured markers indicate the amplitudes at the (off-frequency) CFs one ERB_N below (grey) and one ERB_N above (yellow or orange) the on-frequency CF (502 or 4013 Hz). These markers are highlighted using the same colours in Figure 5. In the third and fourth rows, the average FFT response of two cochlear-filters (CFs of 502 Hz or 4013 Hz) in response to six 500-ms white noise sections are shown in grey, and the corresponding smoothed responses are shown in colour. This type of responses was used to assess the quality factors of Figure 6. The dashed black lines indicate the corresponding −3-dB filter bandwidths. All responses were shifted vertically by the reference gains given in Table 2 (see the text for details). *Literature*: Figure 1A from [76], Figure 2C E from [77], and Figure 2 from [12].

A relevant observation: Keep in mind that the filters differ in their tuning

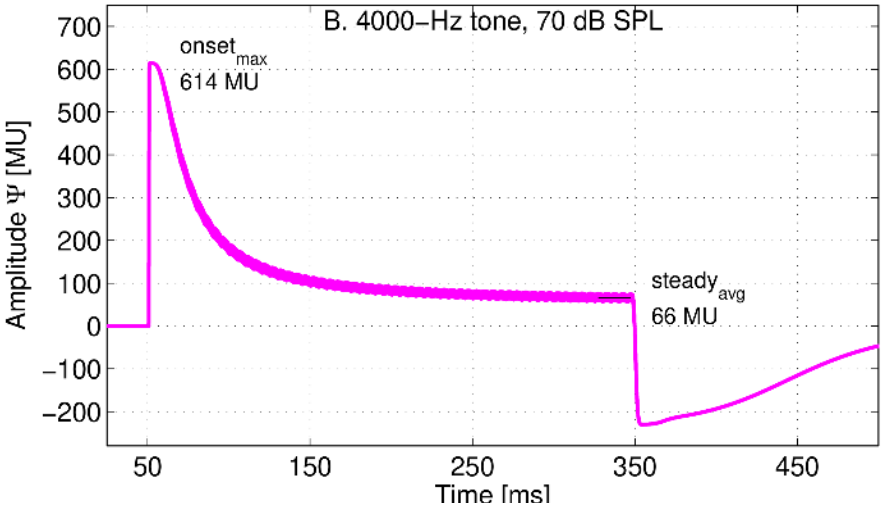
Eight models:

- Biophysical
- Phenomenological
- Functional effective



Stage 5: Auditory adaptation stage

- Learning from observations:



Approximation using “adaptation loops”
(used in dau1997, relanoiborra2019, osses2021)

Fig. C.11B
from Osses (2018, Ph.D. thesis)

Kohlrausch et al. (1992):

“The last class of investigated models tries to incorporate the adaptive properties of the auditory periphery. Adaptation means a change in the transformation characteristic according to the input level. Such an automatic gain control **can be** achieved with a feedback loop […].”

Stage 5: Auditory adaptation stage

- The “more accurate models”:
  - Auditory-nerve synapse models

- These model responses share some qualitative resemblance

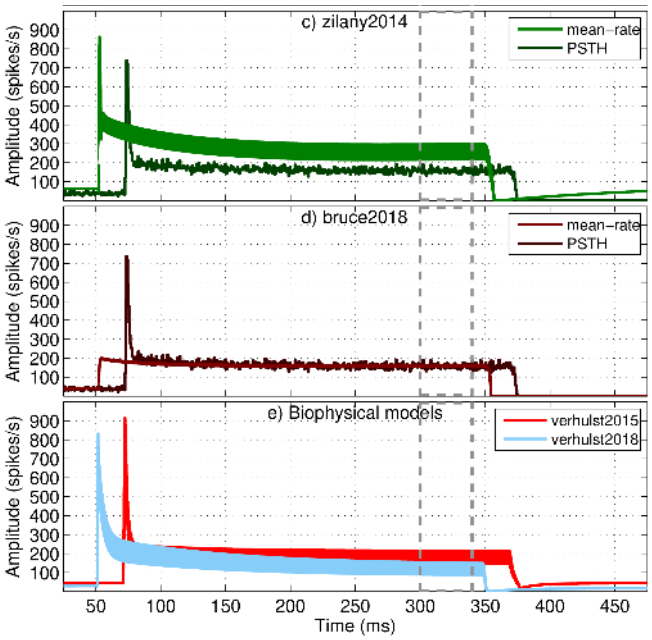


Fig. 9 from Osses et al. (2022)

Synchrony capture

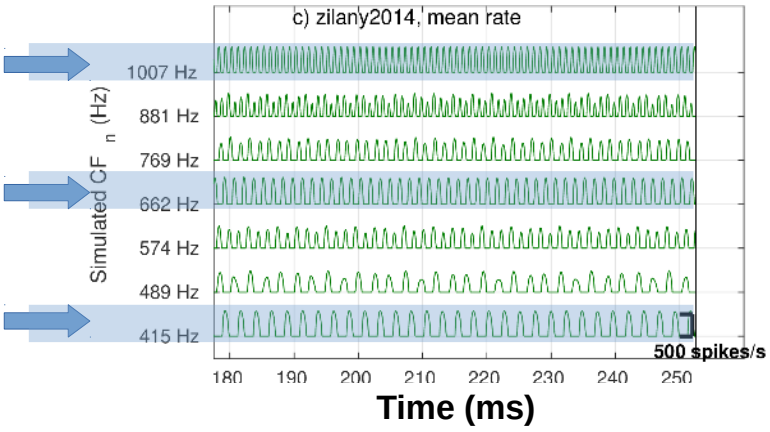
- Occurs when the neural activity in on-frequency channels is driven primarily by one frequency component in the harmonic complex, such are there are minimal fluctuations, while off-frequency channels exhibit fluctuating patterns.

In other words:
- Constant envelope in on-freq. channels
- Variable envelope in off-freq. channels

Synchrony capture – Generating the benchmark

- The stimulus: Three tones at 50 dB SPL but with random initial phase
- How to estimate the envelope of an AN response?
- How to estimate the variability of the envelope?

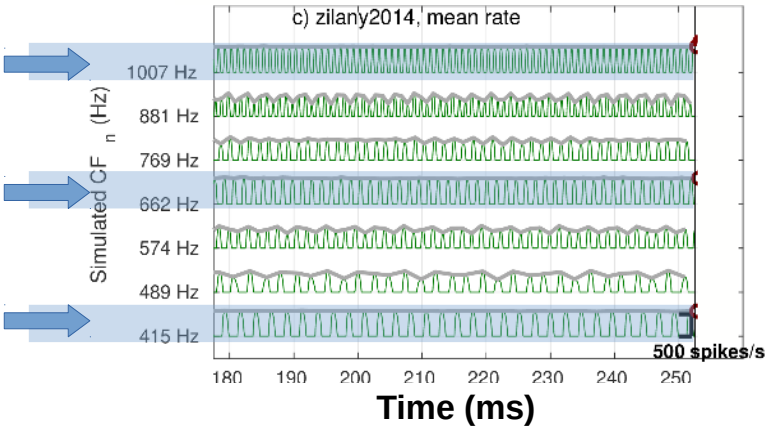
f = 414, 660, 1000 Hz



From Fig. 13

Synchrony capture – Envelope of the AN response

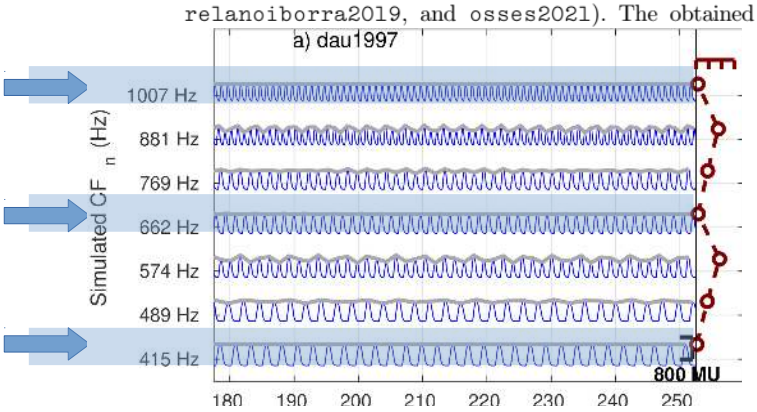
Those envelope fluctuations were constructed by connecting consecutive local maxima that had amplitudes above the mean responses (onset excluded) of each simulated channel.



From Fig. 13

Synchrony capture – Dau et al. (1997) model

Those envelope fluctuations were constructed by connecting consecutive local maxima that had amplitudes above the mean responses (onset excluded) of each simulated channel. Subsequently, the standard deviation of the obtained envelope estimate was (arbitrarily) divided by one thirtieth of the amplitude scales shown in the insets of each panel (e.g., divided by 800/30 MU for `dau1997`, `relanoiborra2019`, and `osses2021`). The obtained



From Fig. 13

Let's get back...

From Maxwell et al. (2020, JASA):

Model by Zilany et al. (2014)

Note that this AN model is substantially different from peripheral models such as Dau *et al.* (1997), used in simulations of the notched-noise task in Derleth and Dau (2000) and similar masking tasks (with some changes to the model) in Jepsen *et al.* (2008). The Dau *et al.* (1997) model, and later modifications, did not include IHC saturation and did not aim to simulate physiological neural fluctuations and capture.

Another model

True

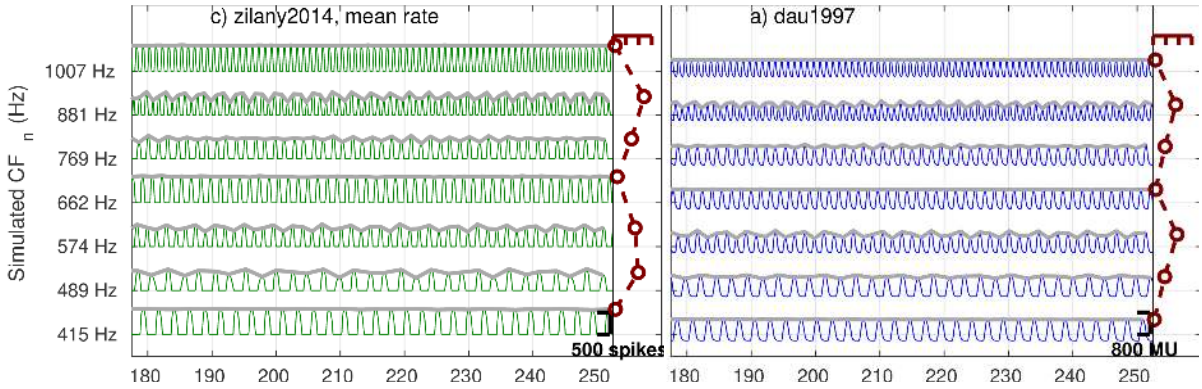
But does IHC saturation really play a role in synchrony capture?

True, but this does not mean that a model may not account for neural fluctuations and capture

- Based on our comparison of models, synchrony capture is more related to the saturation at the AN synapse (or equivalent) model stage
- The main point by Maxwell et al. is that off-frequency channels may contribute signal cues that can be used by listeners in a psychoacoustic task: I fully agree with this...

A summary: Synchrony capture

- This is an interesting effect observed at the **simulated output** of the AN synapse (or equivalent) stage.
- Probably it is difficult to obtain physiological data supporting this (psychoacoustical data by Maxwell)
- **Zilany2014**: It has IHC saturation and an AN synapse model → Accounts for capture
- **Dau1997**: It does not have IHC saturation and it has a different adaptation stage →It seems to still account for capture, **but the model had not been validated for this type of sounds**
- This is based on a metric of "AN envelope fluctuations", but when is this metric significantly different from 0?



A summary: Comparison of models

- This study (the paper) compares different models and use them on the same set of sounds, independent of the rationale of each model
- Different models will give different (but hopefully comparable) results:
  - Make sure you are aware of the capabilities of the specific models
  - **Warning**: Not all model descriptions are always available (codes somewhere?)
- Different models have been validated with different sounds and more complex models are not necessarily best to other models (you need to check that, if not shown in the literature)
- Make sure your simulations are replicable, and that you can use your model with different sets of parameters without having to re-program:
  - For instance, use amt

Questions during the presentation

Question from Ellie: You were presenting about this synchrony capture that seems to be related to on-frequency and off-frequency simulations. Is this also related to the specific choice of simulated CFs that you used in the presentation? For example, in Slides 17-19 you used along the CF axis (the ordinate) a frequency of 1007 Hz instead of 1000 Hz?

This is an interesting question because you are asking whether this phenomenon (capture) is somehow related to the specific configuration of stimulus-model parameters. I mentioned Armin several times during the presentation, and he was also wondering the same thing, he was asking "is it maybe that if the tone frequency falls slightly off frequency (just before or just after the best frequency) that you will obtain different results?." Well, the answer is that I checked these interactions and capture could be accounted independent of the choice of the specific "CF bin", so that 1007 Hz or 1000 Hz would be more and less similarly "captured". The choice of using 1007 Hz instead of "rounded values" was due to the "fair comparison of models". The verhulst2018 and verhulst2015 weren't able to freely simulate just any specific CF but were constrained to 1000 cochlear equidistant points (see Eq. 3 in the paper). To compare across models, we decided to also adopt simulation CFs corresponding to one or more of those 1000 cochlear points. In fact, CF_n=1007 Hz is obtained by using the bin number n=245 in Eq. 3. In the paper, each time we reported a simulated CF we also reported its corresponding bin number. We didn't use this criterion for the frequency of the generated stimuli, and that's why the tones across this paper often have "rounded values" of 1000 Hz or 500 Hz, etc.

Question from Dyan: During the presentation you said that in your comparison you decided to not indicate any "winner model" but was there a clear "looser"?

Well, we decided to refrain from indicating a winner or a looser because a model will be good or bad only when you decide to apply it to a specific problem or application and when it succeeds or fails. Now, all the stimuli we used in this comparison paper were artificially-generated stimuli – tones, clicks, or noises – that sometimes have been used in the development of physiologically inspired models, in contrast to perceptual models, that have focused in the combined use of front-end (up to model Stage 6) with a decision back-end (that I didn't present today). Maybe, if you read the paper, you can find that the model from my group, named king2019.m, seems to be a weak model because it shows some ringing in the frequency response of the cochlear filters in response to white noises (see the third and fourth rows of panels in Fig. 4, in the paper) or that although it includes some adaptation in the simulated responses, it is the only model that doesn't accounts for adaptation saturation. However, the ringing of the model was found in a range of frequencies that its developers didn't use: they tested the model using on-frequency CFs and then +/-2 ERB_N, totalising 5 bands; and furthermore, from all tested models, king2019 is the simplest implementation and the only model that doesn't require extra (python/C/mex) compiled files. So this "disadvantages" might be seen as at cost of implementation efficiency, and I can tell that king2019 it is successful in simulating many AM psychoacoustic experiments for normal-hearing and hearing impaired profiles. In fact, according Torsten Dau – who contributed a very nice paragraph in the discussion section of our paper – any model should be falsifiable or, in other words, it is impossible that a model is perfect. He stated that the limitation of your model should always be made explicit when you publish your model. In my opinion, this helps the reader to get a better impression of reliability… In this line of reasoning, actually once I get to know a new model I first look at the dataset of stimuli used for its validation and then I also look at the reported limitations of the model and ultimately I judge whether I find the model interesting or not...

Question from Wiebe: It looks like these models require some type of stochastic component, especially if you look at the level of the auditory nerve. Many years ago, I had the possibility to do AN recordings in animals. I was primarily a psychophysicist, so that was a great experience, but then I realised how variable are the data you can collect from the auditory nerve. I think you can have a look at some of those data that were published in a paper that I authored in JASA I believe in 1984.

Great comment Wiebe, thank you! And yes! Other discussion points that we had with Armin about these types of plots is that we, as researchers, try to publish beautiful figures, but of course this is not always good, especially if others try to replicate your paper and then they realise that the obtained results are not "as nice as" the figures in the paper. Said this, I think I might have fallen in this same problem: I showed you the outputs of the zilany2014 model using the AN mean firing rate generator using a deterministic configuration: so this is in fact an idealised plot. When I use either this model or bruce2018, if my goal is to be "more realistic" in my simulations, my first preference is to use the random spike generators. In my opinion, the model bruce2018 is very powerful for this. That model uses exactly the same model stages as zilany2014 only different in the ANF synapse model that Ian defined in his 2018 paper.

Other ways to test a stochastic component when your model is deterministic is to try noisy inputs to the model. I find that a first step to have a look at how the model behaves statistically speaking…

Lastly, I want to share one last concern that Armin had about idealised figures. In the context of this synchrony capture, for example, based on what Maxwell et al. showed and what I showed today, I am not sure how much capture would happen if I add noise, say at 0 dB SNR, to the complex tone. It could be that that generates saturation in all on- and off-frequency channels. I can't really tell if I don't test it. I have to emphasise again that what I am showing here is not something I use in my everyday research life and that's why I can give limited comments about what would happen with these stimuli in other contexts, but the main point of this talk is that given a specific definition of a phenomenon (in this case of capture) you can easily check if another model can account for this new phenomenon, even if the model has not been previously validated in such condition(s).

Special thanks to…

Armin, who read all the different drafts of this paper – We actively discussed about synchrony capture and extensively discussed about Figures 4, 5, and 6 of the paper.

References

Dau, T., Kollmeier, B., & Kohlrausch, A. (1997). Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers. *J Acoust Soc Am*, *102*, 2892–2905. doi: 10.1121/1.420344

Barnes, N. (2010). Publish your computer code: It is good enough. Nature, 467(7317), 753. doi: 10.1038/467753a

Kohlrausch, A., Püschel, D., & Alphei, H. (1992). Temporal resolution and modulation analysis in models of the auditory system. In M. Schouten (Ed.), *The auditory processing of speech* (Vol. 10, pp. 85–98). Mouton de Gruyter.

Majdak, P., Hollomey, C., & Baumgartner, R. (2022). AMT 1.x: A toolbox for reproducible research in auditory modeling. Acta Acustica, 6, 19. doi: 10.1051/aacus/2022011

Maxwell, B., Richards, V., & Carney, L. (2020). Neural fluctuation cues for simultaneous notched-noise masking and profile-analysis tasks: Insights from model midbrain responses. J. Acoust. Soc. Am., 147, 3523–3537. doi: 10.1121/10.0001226

Osses, A., Varnet, L., Carney, L., Dau, T., Bruce, I., Verhulst, S., & Majdak, P. (2022). A comparative study of eight human auditory models of monaural processing. Acta Acustica, 6, 17. doi: 10.1051/aacus/2022008

Osses, A., & Kohlrausch, A. (2021). Perceptual similarity between piano notes: Simulations with a template-based perception model. J. Acoust. Soc. Am., 149, 3534–3552. doi: 10.1121/10.0004818

Osses, A. (2018). Prediction of perceptual similarity based on time-domain models of auditory perception. Technische Universiteit Eindhoven. URL: https://research.tue.nl/nl/publications/prediction-of-perceptual-similarity-based-on-time-domain-models-o

Zilany, M., Bruce, I., & Carney, L. (2014). Updated parameters and expanded simulation options for a model of the auditory periphery. *J. Acoust. Soc. Am.*, *135*, 283–286. 10.3389/fphar.2018.00665